Focus

Audiovisual localization

The search for perfection

Kamil Juljanski is a solutions architect at Capita Translation and Interpreting. His main role is to assist the business development and project management teams and use technology to find the perfect solution for their needs.

Kamil Juljanski
When you’re watching TV and you come across an advertisement or a show with an amicable presenter, you will be unlikely to critique their speech. Even if the presenter has a speech impediment or a strong regional accent, and even when they go off script, as long as the intended meaning is conveyed, you accept it for what it is.

However, in the world of localization, we put a lot of emphasis on all of these imperfections due to our traditional workflow and culture. While some audiovisual project requirements are flexible, others are rigid, and will require fixes to the final piece. Content used for voiceovers is not always created primarily to be read aloud and doesn’t always sound natural when spoken, making it more difficult for the voiceover artist. Maybe a word was pronounced slightly differently, or maybe the voiceover artist deviated from the script to improve the flow, but occasionally, the final review requires the localized voiceover to match the source script exactly. In audiovisual localization projects there are many variations to consider: translation of the script, cultural nuances, voiceover artist talent, matching the speed of the target language speech to the source — the list goes on, and if any of these components don’t go according to plan, it impacts the entire project.

With this existing process, the artist might be required to return to the studio and rerecord parts of the script, which then have to be engineered back into the audio. This not only affects the speed of project delivery, but creates additional cost. However, this can be avoided — sound engineers can check if similar words or phrases crop up in other parts of the script, and cut and paste the audio. Using a number of different tools and techniques, it is possible to blend and merge audio, even at the syllable level. Working with pitch and speed allows the studio to add in audio and make it sound natural.

The Photoshop of speech?

But what do you do if the words in question are not repeated elsewhere in the script, or you cannot find the right syllables to slice and dice? Adobe is developing a new tool that could be the answer. Project VoCo was showcased at the Adobe MAX conference in November 2016. This technology allows the user to tell the software which word or short phrase should have been spoken, the software then analyzes the speech already recorded and approximates how that same voice would have spoken the relevant phrase, and recreates the content. In order to work, VoCo requires at least 20 minutes of voice data. It breaks the recording down into phonemes and creates a model of the person’s speech. The user can add or remove phrases by using a simple text editor and the algorithm will edit the recording accordingly.

This is still very much in the development phase, and for now we do not know how, or even if, it will support multiple languages and accents. However, in theory, the accents are not such a problem since the software is analyzing how that person sounds, meaning it is already unique to the user’s voice. The biggest challenge will be multilingual support, as the software currently only supports English. Many languages have unique sounds — for example, the uvular *rrrr* sound, or the throaty *chhhhh* sound in Hebrew.

Nonetheless, the VoCo solution seems to be a good bridge between text-to-speech and real human voiceover, as advanced editing skills allow any fixes to be done quickly, without the original voice actor’s involvement. Using solely text-to-speech sounds too machine-like for long scripts, and even the best systems and most human-like tools require a huge manual effort to program. From a commercial point of view, these tools can sometimes cost more than their human counterparts, including licensing fees for each piece of content.

When producing voiceover or dubbing for movies, additional challenges arise. The translated content not only needs to be accurate, it has to sound natural as well. However, “natural” doesn’t just mean the performance of the voice talent or the quality of translation — good sound design has to transport the audience into the scene and imitate reality. It’s easy to forget that technology has made amazing strides over the last decade because, unlike the visuals, the audience tends to notice audio only when there is a problem with it.

**Sound layering: walla and rhubarb**

To illustrate this point, let’s talk walla. Unless it is a corporate marketing video, or a simple tutorial, most audiovisual content has many layers of sound: music, sound design and even the sound of other people speaking in the background. The background noise of a Japanese office might be different from the noise of an office in London. And this is where walla comes in.

Walla is the sound effect replicating the background noise of a crowd. If you record several people repeating the word “walla,” you can create an auditory illusion of a crowd. But crowds sound different in different languages. In the UK, it’s a fairly well-known joke in the industry that, when speaking in the background, extras repeat the word “rhubarb” to recreate the hubbub of a crowd. Thanks to modern technology and high definition audio, there’s a risk that some of that chatter can now be audible, so to maintain verisimilitude, a group of voiceover actors, also called “walla groups,” record short conversations that are related to the setting. So, a crowd recording a modern script will
have different conversations compared to extras in a medieval story. When translating a dubbed movie, the dialogist needs to write similar conversations in their native language. They need to look at the scene, create characters out of the crowd and think what sort of conversations they might be having. The studio will then record these conversations and mix them together to simulate the sound of the crowd. If any part of the walla dialogue bleeds through and is actually audible to the audience, this no longer poses a problem, but rather helps to set the scene and create atmosphere.

Overcoming challenges in audiovisual projects

Translating for moving images is also far more complex than normal “wild audio.” Wild audio is normal for eLearning or phone prompts, but for a movie or the dubbing of a presenter, it is necessary to synchronize the audio to the timings of the original content. When preparing voiceover content for corporate audiences, the aim is to synchronize each phrase so that the localized audio starts and stops as the speaker does. If the script is too wordy, it is sometimes possible to slow down parts of the video, freeze the frame or even loop parts subtly so it seems the presenter is still speaking while the localized audio completes.

None of the tricks above can help with dramatic content or commercial documentaries, though. These require intricate sound engineering in order to syllable match. The actual production time could be significantly longer too, as the actors would have to deliver the lines multiple times in a variety of speeds. This is a particular challenge when working on animation, as the localized content needs to create an impression of lip-syncing.

The translator needs to map out all plosive (p and b) and nasal (m and n) consonants and make sure that the translation contains same or similar sounds in the same places.

Regardless of the approach and type of content, any problems, whether caused by a mistranslation or the voiceover, become amplified as fixing them is a time-consuming and costly endeavour. Of course, if Adobe can deliver their new technology, it would shape the audiovisual localization process and remove the need to recall or rework the audio. Until this new technology comes to market, how can we, as localization experts, approach such challenges? First, it is important to understand the requirements of the project. What is the approach and who is the audience? How will the audio be synced? Does the length and content of the translation need to be amended? For such intricate tasks, reference materials can be a huge help — glossaries, key names, phrases and pronunciation guides should be prepared and available to all stakeholders, from the translators to the studio.

Technology, expertise and talent

Just as the old proverb states “fail to prepare, prepare to fail,” insufficient planning in audiovisual projects can cause costs to spiral out of control. Delays in sending the translation to the voiceover artist, or the need to correct voiceover in post-production should be accounted for in the overall turnaround time. If the translated script isn’t ready for the day of the recording, the session will have to be rescheduled, adding further costs. Ideally, the translator
will be on-hand during the recording to advise on the translated script, or rephrase certain sentences if they don’t fit or sound quite right. It may sound obvious, but translators should be familiar with the specific requirements of the project, as their translation will shape the voiceover itself. It’s important for the translator to be aware of the length of the text, trying to keep their translation similar in length to the source material and make sure the text is easy to pronounce. Regardless of how good the quality of the translation is, if the voiceover artist stumbles over words, their effort will be in vain. Finally, it’s preferable to have a producer or a director in the studio when recording. They will oversee the process, direct the talent and rectify any issues before the recording session wraps.

Recording voiceover is a challenging, but very satisfying process. We live in an increasingly visual world, reliant on multimedia and dominated by social media and video sharing platforms. YouTube is the second most popular website in the world, 95% of the global population live in an area covered by a mobile network, and two-way communication platforms (such as Alexa and Siri) are becoming commonplace across the globe. Our clients use video and audio to connect with their employees, partners and potential clients, and as such, the demand for high-quality audiovisual localizations grows. In such a subjective and creative industry, technology in the field continues to develop, but there are still challenges to overcome, tongue twisters to untangle and mispronunciations to correct. As long as there is a professional team of translators, voiceover artists, project managers and sound engineers working together behind every project, the quest to achieve audiovisual perfection comes closer into reach. [M]